protected by U.S. Patent 6,640,227 issued 10-28-2003. Unlike all other previously known methods for data clustering, the ETSM-based processing of input data is holistic, i.e. involving the entirety of input data at once; and, therefore, the elaboration of an appropriate procedure for similarity computations has been an extremely important aspect of the ETSM methodology. Moreover, although the principles disclosed in Patent Application 10/622,542 (hereinafter called the '542 application) have been developed as applied to clustering by the ETSM method, they have a more general practical value in various fields of applied sciences and practice which utilize similarity-dissimilarity matrices.

2. Dimensionless basis of similarity computations

2.1. Claim 1 of the '542 application claims a "method for computation of similarity matrices of objects in a high-dimensional space of attributes... on a dimensionless basis". The importance of the 'dimensionless basis' condition is pointed out in sections "Description of the Related Art" (cf. [0014], lines 1-22) and "Detailed Description of the Invention" (cf., for instance, [0049], lines 5-10; [0054], lines 1-13; [0074], lines 8-10; etc.) (these and further references to the '542 application text are based on the layout of Pub. No.: US 2005/0021528 A1), where it is also emphasized that a failure to meet this condition makes theretofore available methods for computation of similarity-dissimilarity matrices both unscientific and contrary to common sense.

2.2. To explain the notion of 'dimensionless basis', let me use a simple example. Assume that objects A and B are described by parameters x and y. It means that in a space of coordinates x and y, objects A and B represent two points. The distance between these two points in the space of coordinates x and y is accepted as the degree of dissimilarity between A and B, whereas the inverse value of thus determined degree of dissimilarity between A and B corresponds to the degree of similarity between A and B. If both parameters, x and y, describe, for instance, the objects' lengths measured in inches, the dissimilarity coefficient has dimensionality of length. If, however, the two parameters describe different properties – e.g. parameter x describing length measured in inches, and parameter y representing time in years (for instance, an animal's life expectancy), the dimensionality of the dissimilarity coefficient appears to combine length (in inches) and time (in years). While a correlation between the length of an animal's body and its life expectancy is known to exist (typically, the lifespan of animals increases with size), i.e. can be established by comparing, individually, the properties described in different dimensions, it is clear the such a dimension as an

"inch-year" is absurd. With high-dimensional spaces of parameters, thus computed similarity coefficients become progressively senseless.

2.3. The foregoing makes it clear that it would be grossly contrary to common sense to attempt the "fusion of different attributes (parameters)" in similarity matrices of objects by measuring the distances between parameters that reflect the objects' properties of physically different nature. Unfortunately, legacy methods not only view it permissible but have actually legalized the use of distances between parameters that are incomparable – for instance, the intensity of color and the weight – ignoring the fact that different properties are described in different units, and – even more importantly – that the scale of quantitative changes is individual for each specific attribute (parameter), i.e. the notions of "extremely high" or "extremely low" values are very different for different attributes. Thus, quantitative measurement of distances between the values of different parameters is completely nonsensical. Nevertheless, the fallacious practice of measuring the distances between parameters with different dimensionality by using Euclidean distance has been widely distributed, and is applied, in particular, by Frederick Herz et al. in U.S. Patent No. 6020883 A1 (hereinafter called Herz) (see, for instance, col 37, lines 12-14, and col 40, lines 47-48).

2.4. The computation of similarity-dissimilarity matrices on the dimensionless basis is possible only by following the procedures defined in Claims 1, 2, 6 and 7 of the '542 application. Indeed, the only way to obtain similarity coefficients on the dimensionless basis is: first of all, to compare the objects according to each parameter individually (Claim 1a) and, second of all, to apply such metrics (see Paragraph 5.3 below) which provide a ratio between the parameter values under comparison, i.e. represent the result of division of one value of the parameter by another value of the same parameter (Claims 6 and 7), which provides the reduction of dimensionality of the set of attributes and produces dimensionless similarity coefficients. Addition, subtraction, or multiplication of parameter values – as it is provided by Euclidean distance, "city-block" metric, or as described by Herz (col 5, lines 55-66, to col 6, lines 1-10) – cannot provide dimensionless computation of similarity coefficients. Euclidean distance is a distance between two points which can be determined by application of the Pythagorean theorem. Unlike our technique for establishing distances between objects by computation of monomer similarity matrices and hybridization thereof, Euclidean distance is measured directly in an n-dimensional space by applying the Pythagorean formula wherein the number of the terms of the equation corresponds to the number of parameters, that number being n.

3. Monomer matrices

3.1. The totality of the procedures according to Claims 1, 2, 6 and 7 of the '542 application provides dimensionless similarity coefficients based on comparison of the objects by each individual parameter, thus resulting in production of a set of similarity matrices for a given set of objects, wherein each similarity matrix reflects the objects' similarities by one parameter – hence we call them monomer similarity matrices. The number of thus obtained monomer similarity matrices corresponds to the number of parameters describing the objects. As monomer similarity matrices are based on similarity coefficients that are dimensionless, they can be easily fused together, i.e. hybridized.

3.2. It must be remembered that the description of our invention clearly demonstrates ([0020]) that monomer similarity matrices provide for analysis of and comparison between individual parameters from the standpoint of their contribution into the similarities or differences between the objects. In the context of these Comments, it is important to emphasize that monomer similarity matrices provided by the invention disclosed in the '542 application serve neither as the means for comparison between objects' profiles nor for obtaining of combined profiles (as, for instance, the combined profiles of Mom and Dad in the evening and the combined profiles of the children in the afternoon, described by Herz, col 5, lines 25-55).

4. Matrix hybridization

4.1. Once a full set of monomer similarity matrices for a given set of objects is obtained (a 'full set' means that there are as many monomer matrices as there are parameters describing the objects under analysis), the monomer matrices need to be consolidated into a united matrix that would thus reflect the similarities or dissimilarities between the objects, based on the totality of all the involved characteristics (parameters). In our invention, disclosed in the '542 application, the consolidation of monomer similarity matrices computed on the dimensionless basis according to the procedures defined by Claims 1, 2, 6 and 7 is called 'hybridization' as this term accurately renders the essence of the disclosed process of unification of monomer matrices into one general similarity matrix covering all the parameters as a whole.

4.2. The averaging of similarity coefficients through computation of their geometric or arithmetic means, as employed in the process of hybridization of monomer similarity matrices

(Claim 2), is necessary because it provides following: when a hybrid matrix is further processed by the ETSM method with the use of XR-metric (Claim 6), distances between the objects can thus be computed on the dimensionless basis, which would be impossible by using traditional methods for computation of distances.

5. Shape and Power

5.1. As each monomer matrix is computed independently from any other monomer matrix for a same set of objects, it is possible: (a) to apply an optimal metric for each parameter (Claim 3), and (b) to change the weights of parameters by changing the share of a respective monomer matrix in the hybrid matrix (Claim 8). Such an approach to changing the weights of parameters in the totality of parameters describing a set of objects is truly correct and valid since the monomer matrices are based on dimensionless similarity coefficients.

5.2. Claim 3, when considered in isolation from Claims 1, 2, and, particularly, 4 and 5 (which certainly would be contrary to the spirit of invention evaluation), may seem to define a self-evident necessity of applying optimal metrics in computation of similarity matrices. However, there are two factors that make Claim 3 nontrivial. Firstly, the application of an optimal metric for each of the attributes (parameters) is possible only when using the method of monomer and hybrid similarity matrices. Secondly, Claims 6 and 7 not only provide two new metrics – for "shape" and "power" – but also, taking into account Claims 4 and 5, postulate that any of theoretically possible attributes ultimately reflects either shape or power.

5.3. In other words, dimensionless similarities can be established by either division of parameter values (in case of Power attributes, regardless of the clustering method) or subtraction of exponential numbers (in case of Shape attributes and only when applying the ETSM clustering method). In case of the latter, the dimensionless basis of similarity computations is provided by the ETSM processing of data, whereas with traditional methods the similarity coefficients computed by subtraction are not dimensionless.

6. The notions and terms

The notions of "monomer similarity matrix", "hybridization of monomer similarity matrices", the concept of two alternative types of metrics ("shape" and "power" metrics), and the determining of the weights of attributes on a dimensionless basis are pioneer terms introduced by us